

# **IFPS Era Forecast Verification: Endorsements, Requirements, and Recommendations**

2 April 2004

## **An IFPS Science Steering Team (ISST) Position Paper**

**Introduction** – The transition into digital forecasting represents a significant change in the way NWS forecasts are produced and provided. No longer are forecasts conveyed through zone-averaged words through seven days, and through a relatively small set of selected city forecasts through three days. Today, NWS forecasters are producing high resolution forecast grids in both space and time out to seven days. These grids contain information that allows customers and partners to extract detailed forecasts to feed countless uses and applications. Inherent to this dramatic shift in the forecast process paradigm is the increased responsibility of forecast accuracy. As such, attaining the best possible accuracy and skill requires the development and use of sophisticated verification tools and techniques that ideally allow for continual improvements to the digital forecasts and their supporting forecast process.

A properly designed verification system provides important information on differences between current forecast performance and established standards of forecast accuracy and skill – which are needed but yet to be determined – and is used to identify deficiencies and strengths of the digital forecast that help determine future program direction and design. Using performance feedback as an avenue to producing more accurate and skilful forecasts will also help reduce inter-side grid discrepancies in the NDFD.

The IFPS Science Steering Team (ISST) recognizes the need and strongly supports immediate development work on such a verification system. Ignoring or diminishing this need as a critical aspect of digital services sends a powerful message to internal and external customers and partners that accuracy is a low priority. This ISST position paper first summarizes the ISST review of the NWS Verification Integrated Work Team (IWT) plan, outlines what we see as the necessary requirements needed to guide initial IFPS verification system development, then provides recommended first steps and initial tasks needed to construct such a system.

These requirements and recommendations have been adopted from and expanded upon those contained in the NWS Western Region SOO/DOH IFPS Whitepaper, and the subsequent SOO/DOH IFPS Whitepaper Implementation paper (copies of each are available at the ISST website located at [http://www.nws.noaa.gov/ost/ifps\\_sst/](http://www.nws.noaa.gov/ost/ifps_sst/)).

**ISST Review of the National Verification IWT’s “Plan for Verification of Forecasts in the National Digital Forecast Database”** – The ISST was tasked to review the IWT’s plan dated 31 December 2003, and prioritize the 21 tasks of the plan. Overall, the ISST feels the plan represents an adequate foundation to support development of an effective verification system for IFPS. The ISST created these five task categories to help summarize system capabilities and requirements: 1) enhancing the existing MDL system,

2) developing a new system with a relational database management system (RDBMS) with GIS capabilities, 3) providing local office real-time access and display, 4) creating and delivering education and training, and 5) development of an accurate analysis system (termed the *Analysis of Record*, or AoR) to be used for grid-to-grid verification. The key ISST endorsements of the plan are the establishment of program leadership necessary to build and maintain the system, the need for grid-to-grid and grid-to-observation verification (with development of the AoR), verification of supporting numerical weather prediction forecasts, and the importance of user training and education. However, notable plan deficiencies are the lack for any development to support real-time forecaster feedback in AWIPS/D2D and the IFPS Graphical Forecast Editor (GFE) that is paralleled by the fact that the proposed RDBMS component is primarily limited to use on the web, and the absence of any specific task to develop an AoR for OCONUS areas.

The IWT's plan also lacks any specific task to establish operational forecast verification requirements, or to establish standards of forecast accuracy and skill. The following section thus attempts to provide a baseline set of requirements – some of which are or are not satisfied by the IWT's plan – that can be used to help steer development of an effective verification system.

**Verification Requirements** – In order to properly build and implement an effective verification system for use in IFPS, it is necessary to first establish system requirements, including establishing associated standards of forecast accuracy and skill. In particular, requirements and standards are needed to satisfy system use in the field and better ensure subsequent impact on any possible improvements needed in the digital forecast process. Completing these steps is absolutely necessary to maintain an efficient and effective development effort – which is critical in our time of limited resources and immediate demand. We also need a system that is functional and effective for years to come as the digital forecast process continues to evolve.

The following requirements are intended to help guide the efforts of the Digital Services Project Office Assessment Team. Although that team is examining additional action plans to assess all aspects of NWS services, including assessing other components of the digital forecast process (e.g., product timeliness, completeness, and consistency), we will focus on requirements and the associated development efforts needed to design a system that assesses forecast accuracy and skill in the digital era.

**Requirement 1: Verification must be an integral component of IFPS.** Developing and implementing a viable verification system for use on NWS forecast products has historically received inadequate attention, especially one that provides responsive feedback to forecasters in the field. Furthermore, other than comparisons made against statistical guidance, no standards of forecast accuracy and skill have been established or used. Early capabilities have evolved into a relatively simple program on AWIPS (AWIPS Era Verification – AEV), while MDL has also developed and tuned a system to access forecast verification through the Internet, though feedback is not considered to be timely. For the most part, however, applications that provide near real-time forecast feedback (e.g., verifying within 12-hr of observation time) have been developed and

supported by field personnel, such as the capability provided by SOOVER and AWIPSTVerify. In addition, verification capabilities of the past and present, while assessing basic accuracies, only provide limited statistical information at a few (usually less than 3-5) locations within each WFO's forecast area. This was considered suitable to the previous forecast delivery system in place. Just over the CONUS alone, whereby a single WFO is responsible for upwards of 55,000 forecast grid points (the number of grid points in the Albuquerque, NM, CWA IFPS grid domain, at a 2.5-km grid point spacing), the digital era requires the agency to restructure its approach to verification and its supporting system design.

At present there is no objective method to properly assess the accuracy and skill of published forecast grids, or the accuracy of numerical weather prediction model forecasts of sensible weather elements used to populate these grids. Such an evolving forecast system requires constant and useful performance feedback in order to make adjustments to the system that can result in more accurate and skillful forecasts. Early efforts to develop a verification system for IFPS, which include MDL's web interface and FSL's Daily Forecast Critique (DFC), represent a good start to building a viable system appropriate for digital forecasting. The ISST, however, feels much more can and should be done to bring proper verification into the digital era.

**Requirement 2: The IFPS verification system should contain both grid-to-grid and grid-to-observation components.** IFPS takes the agency well beyond crude worded forecasts, whereby customers and partners can now retrieve and use specific forecasts for any number of relatively small areas (i.e., individual grid boxes) across an area, a significant portion of which are between standard observations. Thus it is critical that the verification system include a robust gridded component that bases a record of truth on an accurate analysis of all forecast elements. Accurately assessing our skill between observations is only possible through this component, and provides the proper comparison and evaluation of the new system.

Nonetheless, it is also important to verify grid values and the resulting location-specific text products against observations. This, for example, allows comparisons to legacy products (e.g., Coded Cities Forecasts), and to report to customers and partners forecast performance in and near populated areas. (NOTE: Since grid values do not represent a single point, this effort should include comprehensive user education to avoid misinterpretation of verification results.) Besides utilizing the routine set of observations, the system must be flexible to include as many observations as are available within a field office boundary, and be flexible enough for users to group observations by like characteristic (e.g., by zone, by elevation). An effective observation quality control system must be used to ensure that the most valid gridded- and observation-based verification results are realized.

**Requirement 3: The success of an IFPS verification system is critically dependent on an accurate Analysis of Record (AoR).** The true emphasis of the digital verification effort should be on the gridded verification component. A verification system must accommodate the detail and complexity of the system being verified. This aspect of the

system is highly dependent on an accurate AoR. This is not only key to gridded verification, it also feeds other system applications, such as establishing gridded climatology, and developing statistically post-processed forecast guidance such as MDL's Probabilistic Real-time Interpretation of Models and Forecasts (PRIMF). The resolution of this analysis must match the minimum agreed upon grid-spacing being verified, be consistent across all offices from local to national verification programs, and use all available, quality controlled, data sources. This component must be also developed for OCONUS regions, whose forecasters currently lack any use of an agency-supported high-resolution analysis system.

**Requirement 4: The verification system must be available for real-time display in AWIPS/D2D and the IFPS Graphical Forecast Editor (GFE).** In order for forecasters to effectively assess their forecast skills and thereby make subsequent adjustments to the current forecast, they must have verification data available in real-time within AWIPS and displayable within D2D and the GFE. Most importantly and if and when applicable, forecasters should have the option to directly incorporate error grids into their forecast grid generation to, for instance, apply bias corrections. Learning is also generally enhanced when performance feedback is received in a timely fashion, thus the need to be able to see this feedback as soon as it can be computed. This display and use requirement should complement, and not replace, any necessary capabilities to retrieve verification information via the Internet, such as what's provided with a RDBMS.

**Requirement 5: The verification system should provide comprehensive forecast feedback with flexibility through multiple techniques.** Historically, NWS forecast verification systems have been primarily measures-oriented, such as calculating mean absolute error (MAE) and bias. Though these measures have their value and should be included in an IFPS verification system, their use alone can lead to misinterpretation of verification results and risks missing a complete picture of performance, accuracy, and skill. It says nothing about the forecaster's ability to capture significant changes to the weather (e.g., cold outbreaks), or to capture the general weather patterns. Therefore, to supplement relatively straightforward statistical measures, the IFPS verification system should also employ joint distribution calculations, as well as object-oriented techniques. Distribution-oriented verification allows a system to reveal areas *when or where* forecast skill is particularly weak or strong. Object-oriented verification, or what has also been termed a spatial data mining technique, rewards performance for representative forecasts that are phase shifted. Object-oriented verification is especially useful in verifying forecast positions of tropical weather systems.

The challenge inherent to creating a database so rich in information is developing a delivery and display system that communicates this information to the field and to decision makers in an effective and efficient manner. Thus it is imperative that a verification system include a useful set of display tools to view performance feedback in a variety of ways and complexities, from simple "quick-look" views, to those that may require more time to derive.

If a system is intended to be operated at multiple (local to national) levels, each should be consistent in its design to better ensure similar performance feedback interpretations. Internal users of the verification system should be able to extract verification information in multiple ways, whereby results can be derived to address various forecast aspects (e.g., verification of all grid locations above 8000 ft). Such a system likely will require development and use of more sophisticated data management tools, such as those available with a RDBMS. Finally, the accuracy of numerical and statistical weather prediction forecast elements – down to those derived from “Smart Init” routines and displayable in GFE – should be assessed in the verification system, along with the ability to use other benchmark (and typically less skilled) forecast techniques, like climatology and persistence (for measure-of-improvement calculations).

**Requirement 6: The verification system should include a comprehensive training and education program.** Educating system users is critical to helping ensure its success and greatest impact on performance improvement. This includes educating internal and external users. Even relatively simple statistical measures like MAE and bias can be misinterpreted if used carelessly. More complex distribution and object-oriented methods will require training and education to properly interpret results. Education of external users is especially important to ensure correct interpretation of a gridded system where forecasts necessarily represent grid box averages and contrasts to a point based system. This also applies to educating users on the definition of the underlying verifying analysis so as to not misuse grid and observation-based verification results. Furthermore, such misinterpretation can lead to unrealistic expectations of a gridded system providing point accuracy.

**Requirement 7: A select set of verification results and standards for forecast accuracy and skill should be made available to all customers and partners.** Any verification system is highly dependent upon establishing standards of forecast accuracy and skill. These standards are critical to identifying forecast goals that will help determine system design and possible modifications. Making certain that verification results and system development plans are available to customers and partners provides them with a clear understanding of system goals, possible deficiencies, and steps proposed to reduce those deficiencies, or to sustain and exceed these goals. Customers and partners should also be provided guidance and tools to calculate measures of forecast value, with a flexibility to calculate measures applicable to their use of the digital forecasts.

**Recommendations** – In order to satisfy these requirements as quickly, efficiently and effectively as possible, the ISST has identified a set of first steps and tasks needed to get a viable verification system in place for IFPS. It is understood that limited resources exist to complete these tasks and that compromises may need to be made. Yet, this should not reduce their importance in any way. Rather, this absolutely necessitates the elimination of redundant efforts and requires careful collaboration and coordination among current and future development and program efforts. Resources must be leveraged among universities and research and development laboratories to maximize

productivity and to follow an effective implementation timeline. As such, program leadership – both initially and ongoing – is critical to the ultimate success of the system.

**Recommendation 1: The newly formed Assessment Team of the Digital Services Project Office is supported by the ISST to develop an implementation plan that brings a comprehensive verification system and initial operating capability to fruition.** Specifically, this team should focus on formulating an implementation plan to construct a viable verification system based on the proposed baseline requirements previously stated in this position paper. A critical outcome of this team must be to establish an action plan to develop standards for what constitutes acceptable forecast accuracy and skill. Without these standards, any use of verification will be incomplete and greatly ineffective. These standards will also help convey forecast quality goals to external customers and partners who will be investing resources on the use of a forecast system that promises to continue to evolve and improve. Establishing these standards will not be a trivial task, especially to gauge performance in the uncharted area of producing deterministic seven day forecasts. As a first step, climatology and persistence grids, along with statistical and numerical weather prediction model forecast grids, could be used as benchmarks to compare published forecast grids against. Official NWS forecasts should, at a minimum, provide improvement over these methods. Another skill measure is to use a previous forecast performance level measured at a time prior to the implementation of digital services as a baseline improvement value.

This team should carefully consider the existing verification capabilities in place as early building blocks, including enhancing MDL's web based system to include a more representative gridded verification system, providing verification results in D2D and the GFE, and providing more immediate forecast feedback. Existing verification display capabilities in GFE should also be considered. Namely, FSL's Daily Forecast Critique should be strongly considered as providing a viable framework for verification use in the GFE.

Once the work of this team has been completed, the ISST proposes formation of a permanent team or group to oversee implementation, operation and maintenance of the resulting verification system. This team should specifically be tasked to revisit and (when deemed necessary) refine verification system requirements of the agency, ensure adequate funding for its operation, ensure consistency across all verification programs, and make certain that other ongoing and future development efforts that contribute to meeting these requirements are identified and accelerated through adequate support. This team should establish a broad base by tapping expertise external to the agency as peer reviewers (e.g., universities and research laboratories). Team membership should include field experts and represent all NWS Regions, NCEP, and NWSHQ (OS/OST/OH).

**Recommendation 2: Begin immediate work on developing an accurate Analysis of Record for both CONUS and OCONUS.** In order to develop an AoR, the NWS needs to first investigate what research and operational analysis systems are currently available, as well as facilitate dialogues with internal and external experts on this issue. To begin

this process, the ISST has recommended holding an AoR “summit” sponsored by the USWRP. Findings from this workshop, which will likely take place in the summer of 2004, should be used as a framework for AoR development. To help address gridded verification requirements locally, regionally, and nationally, a prototype gridded verification project should be established in the meantime that uses existing analysis systems already in place. A strong candidate to use in this project is the high-resolution ARPS Data Analysis System (ADAS) produced and operated by the Cooperative Institute for Regional Prediction (CIRP) at the University of Utah. Possible efforts could include either expanding ADAS nationally from its current domain over the western two-thirds of the U.S., or developing a capability to locally run a simplified ADAS at each WFO. Additional observations to satisfy a national expansion could come from FSL’s Meteorological Data Assimilation Ingest System (MADIS). Quality control issues may be the principle obstacle to overcome in a national expansion of ADAS, or in any other national system deployed. All AoR work should involve the experience and knowledge provided by the NCEP/EMC, and should also involve and coordinate with distributive hydrologic modeling work in progress.

Since OCONUS areas are typically very data-sparse, development of an OCONUS analysis system must include use of quality controlled remote sensing data. Examples of useable data sources include QuikSCAT, GOES sounder data, and the polar orbiting satellites.

**Recommendation 3: Design and implement a Prototype Verification Project to develop a real-time verification system for WFOs and within AWIPS D2D/GFE.**

This project, modeled off the success of the IFPS Rapid Prototype Project (RPP), should involve a subset of WFOs to test and provide feedback on early system versions and capabilities. The project could use an enhanced version of the MDL system to display and interrogate, in real time, verification data and grids within AWIPS D2D/GFE. Another possibility could be to use other existing analysis systems run locally as previously mentioned. This latter option would mitigate data transmission issues associated with sending high-resolutions verification grids from a centralized location.

Verification data (e.g., database systems) not viewable on these platforms should also be made available, but should not be considered a replacement to any D2D or GFE display capability. An important feature of this system must be flexibility. This would include the capability to derive a variety of verification statistics by region, office, and ideally by forecaster (NOTE: Though individual forecast ownership is deemphasized in the digital process, the need still exists to provide individual forecaster feedback.). The observation verification component should be expanded to include additional observations available from various mesonets. Functionality of the prototype verification system should be blended in some fashion with the existing GFE DFC framework.

**Recommendation 4: Promote research on, and potential development of, a more robust verification system than currently exists.** While the ISST promotes emphasis on a GFE-based verification application to provide essential forecaster feedback in real-time, certain platform characteristics would limit system flexibility if used by itself. Though

current web-based verification feedback capabilities in place provide a good foundation, the agency should promote and support research and development on a more robust web-based system that uses RDBMS and GIS technologies. The goal of this system would be greatly enhanced system flexibility by allowing a wide array of possible performance feedback. Again, any RDBMS shall complement, not replace, any AWIPS/D2D and GFE functionality.

**Summary** – IFPS represents an important and necessary shift in the way NWS forecasts are produced and used by NWS customers and partners. In order to assess accuracy and skill in the digital forecasts, the ISST urges that a more comprehensive verification system be developed and implemented as soon as conceivably possible. Key to this effort is the development of a gridded verification component. This effort relies heavily on the development of an AoR that must optimally match a minimum agreed upon grid spacing. Without this key system component, this verification system will not achieve what it sets out to do – produce an objective and scientific method to measure forecaster skill in producing gridded forecasts. This ISST position paper puts forth a set of what is believed to be essential baseline requirements and associated recommendations to begin to formulate an effective performance evaluation system. The ISST strongly believes that the development of a robust and effective verification system that efficiently delivers performance feedback information, and the continued research into new verification methods, will ultimately help keep the agency at the forefront of weather forecasting services.